An Approach for Finding Saliency Regions in 3D Images and Videos

Joby Anu Mathew
Student (Mtech) ,
Caarmel EngineeringCollege
MG University
Pathanamthitta,India

Salitha M K
Asst.Prof,
Caarmel Engineering College
MG University
Pathanamthitta,India

Abstract-Multimedia processing applications need newly improved techniques for satisfying the new demands in the modern era. 3D multimedia applications are the new trend in the society. The main difference between the 3D and the 2D images is the depth. In addition to the color, luminance and texture , the depth factor is the major feature of stereoscopic display. So, the saliency detection models for 2D display cannot be used for stereoscopic visuals. In this paper, a simple approach is proposed for both stereoscopic images and videos. Features like color, luminance, texture and depth are extracted from the original image and convert to YCbCr color space and also from the discrete cosine transform coefficients of the feature's. Then, a Gaussian model of the spatial distance between image patches a gradient filter is used for calculating the depth map. From that the feature map is constructed and all the feature map are combined to produce the final saliency map for 3D images and videos.

Keywords-Stereoscopic images, Stereoscopic saliency detection, center bias factor, human visual acuity.

I. INTRODUCTION

The saliency regions are the most important or the most noticeable region in an image. The saliency tries to mimic how a human eye identify important objects of the scene and is typically based on a simple fundamental that is the contrast between an object and an neighbor. For visual information processing, HVS(Human Visual System) is an important characteristic. The vision can be broadly classified into monocular and binocular vision. Each serve a unique purpose. The difference between the two is the ability to judge distances or have depth perception. Monocular vision is seeing with only one eye at a time. When both eyes are used, this would become binocular vision. In binocular vision, two eyes work together to focus on a single point. It processes that information to determine depth or distance to that point. Thus, binocular vision is used to determine the depth feature of a stereoscopic image. This is sometimes referred as binocular disparity. In other words, it refers to the difference in image location of an object seen by the left and right eyes, resulting from the eyes horizontal separation (parallax).

In computer vision, binocular disparity is calculated from stereo images taken from a set of stereo cameras. The variable distance between these cameras, called the baseline can affect the disparity of a specific point on their respective image plane. However in computer vision, binocular disparity is referred as coordinate differences of the point between the left and right images instead of a visual angle. The units are measured in pixels. Thus, binocular disparity helps to find the depth perception effectively. The other features are obtained from DCT coefficients of image patches. Discrete Cosine Transform (DCT) is a powerful transform to extract the features. Features like color, luminance, and texture are extracted from the DCT coefficients. The depth saliency is calculated and a Gaussian model is also applied to obtaining the feature map. Then, fusing all the feature map helps in construction of the final saliency map.

The saliency detection has numerous applications. One such application is salient object segmentation. Content aware re-targeting, visual quality assessment, visual coding, 3D video coding,3D rendering etc. are the some of the applications of the saliency detection. The saliency region, that is automatically detected depends on many different applications in image processing. For example, the saliency detection is used for image compression to encode saliency regions in high quality and to increase the compression rate for non – salient regions. Another example, the automatic production of short video is summarized by selecting important shots and scenes from a video.

II. RELATED WORK

Visual attention mechanism has two types, bottom-up and top-down. Bottom-up approach is an apperception process for selecting the salient regions automatically in natural scenes. Top-down approach is a cognitive task-dependent process affected by the performed task. For 2D multimedia applications Jonathan Harel proposed Graph Based Visual Saliency(GBVS) [2].GBVS consist of two steps mainly, forming the activation map on certain feature channels and normalizing it by highlighting conspicuity and admits combination with other maps. Another model by Xiaodi Hou and Liqing Zhang proposed a simple method for the visual saliency detection [4]. This model is independent of features, categories or other forms of prior information of the objects. In this, it first analyze the log spectrum of an input image and then extract the spectral residual of an image in spectral domain and then proposes a fast method to construct the corresponding saliency map in spatial domain. Based on this model

Chenlei Guo and Liming Zhang proposed a saliency detection algorithm based on the phase spectrum, in which saliency map is calculated by the Inverse Fourier Transform on a constant amplitude spectrum and the original phase spectrum [14].

Christel Chamaret and his colleagues made a study on the problems of the 3D processing like disparity management and the impact in viewing 3D scene on stereoscopic screens. In this, the 3D experience is improved by applying some effects related to ROI [20]. Potapova introduced a 3D saliency detection model for robotics task by incorporating the top-down cues into the bottom-up saliency detection [12]. Later, Wang proposed a computational model of visual attention for the 3D images by extending the traditional 2D saliency detection methods. In this [13], the authors provided a public database with the ground-truth of eye-tracking data.

From the above studies, the 2D saliency detection makes use of features like color, luminance and texture only. For the 3D saliency detection depth is the major feature. Thus a simple approach is proposed for both the 3DS images and videos, by taking the depth feature into account.
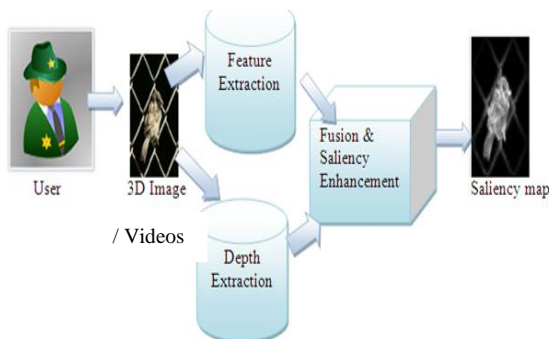
III.SYSTEM MODEL

Fig 1.System model

The system model is depicted in Fig.1.The image is given as the input and color, luminance, texture are extracted from the left and right images.The depth map is constructed from the difference of disparities between the left and right images. Then the feature map is constructed from the depth map. By usingall the feature map the final saliency is constructed. The system consists of threephases.

A. Feature Extraction

This phase consist of three steps. They are (1)Conversion to YCbCr color space, (2) DCT calculation , (3) Depth map calculation.

YCbCr is a family of color spaces and is used as a part of color pipeline in video and digital photography system. The first step is performed to extract the feature like color, texture and luminance.. The Y represents the luminance, Cb and Cr are the two color components. The given RGB image is converted to YCbCr color space. Then in the next step, the DCT coefficient's of YCbCr color space is calculated. The DCT coefficients give the feature like colour, texture and luminance. The DCT coefficients

of Y gives the luminance feature, the DC coefficients of Cb and Cr gives the color feature and the texture feature is obtained from the AC coefficients of the Y component. Finally the depth feature is calculated in this phase. The left and right image of the given image are converted to gray scale. The disparity between the left and right image is calculated and it is slide across each other to get the high confidence disparity map. Then calculate the CSAD (Cost of Sum of Absolute Difference) and CGRAD (Cost of Gradient of Absolute Difference). A gradient filter is used to extract the feature signatures. Then final depth map is calculated by checking the noise in the disparities and checking the boundary to ensure that the disparities are correctly lined up.

B. Depth Saliency Calculation

The depth map is taken as the input for feature map calculation. Firstly divide the depth map into 8 x 8 blocks and obtain the DC coefficients of the image. Then calculate the distance between the image patches and compute the Gaussian value. The Gaussian value is calculated by,

$$\text{Csf} = \frac{1}{g\sqrt{2}} * \frac{dist(i,j)^2}{2g^2} \qquad (1)$$

where *Csf* represents the Gaussian value, *dist(i,j)* represents the spatial distance between the image patches *i* and *j* and *g* is the Gaussian kernel parameter and which is set as 20.The D_{sal} depth saliency is calculated by the rc_{diff} and Eqn.1. That is

$$D_{sal} = \sum(\sum rc_{diff} * \text{Csf}) \qquad (2)$$

where rc_{diff} is the absolute change in the dc coefficients of the image. Then , after normalizing the Eqn 2. , the depth saliency is obtained.

C. Saliency Estimation From feature Map fusion

The feature maps is calculated by Eq. (1).That is the the feature maps of luminance(y_{sal})and the two color components are found by,

$$\text{Cr}_{sal} = \sum\left(\sum\left(Cr_{diff} * CSf\right)\right) \qquad (3)$$

$$\text{Cb}_{sal} = \sum\sum\left(Cb_{diff} * Csf\right) \qquad (4)$$

$$\text{Y}_{sal} = \sum\sum\left(y_{diff} * Csf\right) \qquad (5)$$

where the Cr_{sal}, Cb_{sal} and the y_{sal} are the feature maps of the two color components and luminance. The $Cr_{diff}$is the absolute change in the Cr, the $Cb_{diff}$is the absolute change in the Cb and the y_{diff} is the absolute change in the y component. Then the final saliency is calculated by fusing Eqn.3, Eqn.4 and Eqn.5.That is,

$$\text{Final}_{sal} = \frac{(y_{sal} + Cb_{sal} + Cr_{sal} + D_{sal})}{4} \qquad (6).$$

Then the final saliency is enhanced by applying the center bias factor.

IV. CONCLUSION

An approach for finding the saliency region in 3D images and videos is proposed. The features like color, luminance, texture are extracted by converting RGB to YCbCr color space and depth are extracted from disparities between the left and right images. Then the depth saliency is estimated based on the energy contrast weighted by a Gaussian model and spatial distances between image patches. From the depth saliency the feature maps are calculated and then fusing all the feature map the final saliency is constructed.The proposed saliency detection enhances the stereoscopic applications and it is quite simple to implement.

REFERENCES

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[2] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. NIPS*, 2006, pp. 545–552.

[3] N. D. Bruce and J. K. Tsotsos, "Saliency based on information maximization," in *Proc. Adv. NIPS*, 2006, pp. 155–162.

[4] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007,pp. 1–8.

[5] Y. Fang, Z. Chen, W. Lin, and C.-W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. ImageProcess.*, vol. 21, no. 9, pp. 3888–3901, Sep. 2012.

[6] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Salient region detection by modeling distributions of color and orientation," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 892–905, Aug. 2009.

[7] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*,Jun. 2010

[8] J. Yan, J. Liu, Y. Li, Z. Niu, and Y. Liu, "Visual saliency detection via rank-sparsity decomposition," in *Proc. IEEE 17th ICIP*, Sep. 2010, pp. 1089–1092.

[9] Z. Lu,W. Lin, X. Yang, E. Ong, and S. Yao, "Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1928–1942, Nov. 2005.

[10] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes:The role of global features in object search," *Psychol. Rev.*, vol. 113,no. 4, pp. 766–786, 2006.

[11] Y. Fang, W. Lin, C. T. Lau, and B.-S. Lee, "A visual attention modelcombining top-down and bottom-up mechanisms for salient object detection," in *Proc. IEEE ICASSP*, May 2011, pp. 1293–1296.

[12] E. Potapova, M. Zillich, and M. Vincze, "Learning what matters: Combining probabilistic models of 2D and 3D saliency cues," in Proc. 8th Int. Comput. Vis. Syst., 2011, pp. 132–142.

[13] J. Wang, M. Perreira Da Silva, P. Le Callet, and V. Ricordel, "Computational model of stereoscopic 3D visual saliency," IEEE Trans. Image Process., vol. 22, no. 6, pp. 2151–2165, Jun. 2013.

[14] C. Guo and L. Zhang, "A novel multi-resolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.

[15] A. Treisman and G. Gelade, "A feature-integration theory of attention,"*Cognitive Psychol.*, vol. 12, no. 1, pp. 97–136, 1980.

[16] J. M. Wolfe, "Guided search 2.0: A revised model of visual search," *Psychonomic Bull. Rev.*, vol. 1, no. 2, pp. 202–238, 1994.

[17] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deploymentof visual attention and how do they do it?" *Nature Rev., Neurosci.*, vol. 5,no. 6.

[18] N. Bruce and J. Tsotsos, "An attentional framework for stereo vision,"in *Proc. 2nd IEEE Canadian Conf. Comput. Robot Vis.*, May 2005.

[19] Y. Zhang, G. Jiang, M. Yu, and K. Chen, "Stereoscopic visual attentionmodel for 3d video," in *Proc. 16th Int. Conf. Adv. Multimedia Model.*2010

[20] C. Chamaret, S. Godeffroy, P. Lopez, and O. Le Meur, "Adaptive 3Drendering based on region-of-interest," *Proc. SPIE*, vol. 7524, Feb. 2010.